AD-A094 066 MASSACHUSETTS INST OF TECH LEXINGTON LINCOLN LAB F/G 12/1
SOME REMARKS ON THE GAUSSIAN DISCRIMINANT.(U)
OCT 80 L K JONES F19628-80-C-0002
UNCLASSIFIED TN-1980-47 ESD-TR-80-193 NL

1 OF 1
AD A094066

END
DATE
FILMED
2-81
DTIC

AD A0[illegible]066

LEVEL

Technical Note

[illegible]

[illegible] Jones

DTIC
ELECTE
S
JAN [illegible] 1981
D

[illegible] October 1980

Some Remarks on the Gaussian Discriminant

Prepared for the Department of the Air [illegible]
under Electronic Systems Division Contract F[illegible] by

Lincoln Laboratory

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

LEXINGTON, MASSACHUSETTS

Approved for public release; distribution [illegible]

FILE COPY

81 1 [illegible]

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

LINCOLN LABORATORY

SOME REMARKS ON THE GAUSSIAN DISCRIMINANT

L. K. JONES

Group 92

TECHNICAL NOTE 1980-47

3 OCTOBER 1980

LEXINGTON MASSACHUSETTS

ABSTRACT

We comment on the performance of the Gaussian discriminant function with (possibly) non-Gaussian underlying distributions. An asymptotic expression for the probability of error for the Gaussian case is given with a formal convergence proof.

I. INTRODUCTION

For many practical problems in two class pattern recognition, one has (reliable estimates of) the first two moments of each class (mean vectors in R^n - M_1, M_2 and covariance matrices -Σ_1, Σ_2). Whether or not the underlying distributions are indeed Gaussian, one proceeds to apply the standard Gaussian hypothesis test to classify new data. More precisely, one uses the Gaussian discriminant function $h(X) = \log \frac{p_2(X)}{p_1(X)}$, where p_1, p_2 are multivariate normal with the same first two moments as the underlying distributions. Applying an affine transformation to our problem (which has no effect on the discriminant h) that simultaneously diagonalizes Σ_1 and Σ_2 ($\Sigma_1 \to I$, $\Sigma_2 \to \Lambda$, $M_2 \to \vec{0}$, $M_1 \to (d_1, d_2, \ldots, d_n)$ with $d_\ell \geq 0$), we have

$$(1) \quad h(X) = \tfrac{1}{2} \sum_{\ell=1}^{n} \left[(x_\ell - d_\ell)^2 - x_\ell^2/\lambda_\ell + \ln(1/\lambda_\ell) \right]$$

In this correspondence, we first present some elementary inequalities in h, valid regardless of the class distributions; and then we demonstrate the asymptotic result:

$$(2) \quad P_{error} \approx \frac{1}{\sqrt{2\pi}} \int_{\frac{\sqrt{J}}{2}}^{\infty} e^{-\frac{1}{2}x^2} \, dx \quad \left(\begin{array}{l}\text{with } J \text{ the divergence} \\ \text{of } p_1,\ p_2\end{array}\right)$$

for the case of equal priors, Gaussian distributions, and all λ_ℓ close to 1. We note that the above does not follow from the elementary fact that, for fixed n, $h(X) \to$ a linear function as all $\lambda_\ell \to 1$; for all λ_ℓ may be close to 1 but the quadratic part of $h = \frac{1}{2}\sum_1^n x_\ell^2 \left(1-\frac{1}{\lambda_\ell}\right)$ may not approach 0 if n becomes large.

II. THE GAUSSIAN DISCRIMINANT FOR ARBITRARY CLASS DISTRIBUTIONS

Calculating the first moments of h under each hypothesis, we have, regardless of the underlying distributions:

$$(3) \quad E_1(h) = \frac{1}{2}\sum_{\ell=1}^{n}\left[\left(1-\frac{1}{\lambda_\ell}\right) - \frac{d_\ell^2}{\lambda_\ell} + \ln(1/\lambda_\ell)\right]$$

$$(4) \quad E_2(h) = \frac{1}{2}\sum_{\ell=1}^{n}\left[(\lambda_\ell - 1) + d_\ell^2 + \ln(1/\lambda_\ell)\right]$$

Since $Z-1 + \ln(1/Z) \geq 0$ for all $Z > 0$, we see immediately that

$$(5) \quad E_2(h) \geq \frac{1}{2}\sum_1^n d_\ell^2 = \frac{1}{2} D^2 .$$

Noting that the maximum value of $f(Z) = 1-\frac{1}{Z} - \frac{\gamma^2}{Z} + \ln(1/Z)$ for $Z > 0$ occurs at $Z = 1+\gamma^2$, we have $f(Z) \leq 1 - \frac{1}{1+\gamma^2} + \ln\left(\frac{1}{1+\gamma^2}\right) - \frac{\gamma^2}{1+\gamma^2} \leq -\frac{\gamma^2}{1+\gamma^2}$. Hence

(6) $E_1(h) \le -\frac{1}{2} \sum_{1}^{n} d_\ell^2/1+d_\ell^2$

which is $\approx -\frac{1}{2}D^2$ if each component d_ℓ is small. Therefore, in many practical problems $E_2(h) - E_1(h) \gtrsim D^2 = \sum_{1}^{n} d_\ell^2$. D^2 is then a first order measure of the performance of h. If n is large, the λ_ℓ are close to one, the d_ℓ are small, and the sequence of random variables x_ℓ is k dependent for small k, then we could apply the central limit theorem and obtain estimates of the error probability of h by calculating $VAr_1(h)$ and $VAr_2(h)$ from sample data.

III. ASYMPTOTIC APPROXIMATION TO ERROR PROBABILITY

To justify the claim in I, we state and prove the following theorem:

Theorem : Let a sequence of decision problems, with underlying Gaussian distributions described by means D^i, $\vec{0}$ in R^{n_i} and covariances I, Λ^i, be given. Then, if $\max_{i \le \ell \le n_i} |\lambda_\ell^i - 1| \to 0$ as $i \to \infty$,

$$\left| P^i_{error} - \frac{1}{\sqrt{2\pi}} \int_{\frac{\sqrt{J^i}}{2}} e^{-\frac{1}{2}x^2}\, dx \right| \to 0$$

for the equal prior case.

Proof: We shall apply a central limit theorem for arrays of random variables and use the first two moments of h^i to obtain an asymptotic expression for the error probability. Calculating the variances under each hypothesis of h^i, we obtain

$$(7)\quad \mathrm{VAr}_1(h^i) = \tfrac{1}{2}\sum_{1}^{n_i}\left[\left(1-\frac{1}{\lambda_\ell^i}\right)^2 + \frac{2\left(d_\ell^i\right)^2}{\lambda_\ell}\right]$$

$$(8)\quad \mathrm{VAr}_2(h^i) = \tfrac{1}{2}\sum_{1}^{n_i}\left[\left(\lambda_\ell^i-1\right)^2 + 2\lambda_\ell^i\left(d_\ell^i\right)^2\right]$$

Using (3), (4), (7) and (8), and noting by elementary calculus that

$$\frac{\left(1-1/\lambda_\ell^i\right)^2}{1-1/\lambda_\ell^i + \ln\left(1/\lambda_\ell^i\right)} \longrightarrow \frac{-2\left(1-1/\lambda_\ell^i\right)}{\left(\lambda_\ell^i-1\right)} \longrightarrow -2$$

$$\frac{\left(\lambda_\ell^i-1\right)^2}{\lambda_\ell^i-1 + \ln\left(1/\lambda_\ell^i\right)} \longrightarrow \frac{2\left(\lambda_\ell^i-1\right)}{\left(1-\lambda_\ell^i\right)} = +2$$

$$\frac{2\left(d_\ell^i\right)^2}{\lambda_\ell^i} \bigg/ \frac{-\left(d_\ell^i\right)^2}{\lambda_\ell^i} = -2$$

$$\frac{2\lambda_\ell^i\left(d_\ell^i\right)^2}{\left(d_\ell^i\right)^2} \longrightarrow +2\ ,$$

we have

$$\mathrm{VAr}_1(h^i) \Big/ 2E_1(h^i) \longrightarrow -1$$

and $$\mathrm{VAr}_2(h^i) \Big/ 2E_2(h^i) \longrightarrow +1 \quad .$$

Futhermore

$$-\left(d_\ell^i\right)^2 \Big/ \frac{\left(d_\ell^i\right)^2}{\lambda_\ell^i} \longrightarrow -1$$

and $$\frac{\left(\lambda_\ell^i - 1\right) + \ln\left(1/\lambda_\ell^i\right)}{\left(1 - \frac{1}{\lambda_\ell^i}\right) + \ln\left(1/\lambda_\ell^i\right)} \longrightarrow \frac{1 - 1/\lambda_\ell^i}{1/\left(\lambda_\ell^i\right)^2 - 1/\lambda_\ell^i} \longrightarrow -1$$

imply that

$$\frac{E_2(h^i)}{E_1(h^i)} \longrightarrow -1$$

or equivalently

$$E_2(h^i)/J^i \longrightarrow +\tfrac{1}{2}$$

$$E_1(h^i)/J^i \longrightarrow -\tfrac{1}{2} \quad .$$

We now proceed with the main proof. We may assume (by passing to subsequences if necessary) that both J^i and P^i_{error} are convergent sequences (possibly to $+\infty$ in the case of J^i). We divide the argument into several cases:

CASE (1) $J^i \to 0$

It suffices to show that $P^i_{error} \to \frac{1}{2}$. This is actually true in general. Consider any 2 positive density functions, p, q, on some probability space. Then, if for some real $\delta > 0$, there is no measureable set whose q measure is greater than δ and such that on this set $q/p > 1+\delta$, it follows that

$$P_{error} = \tfrac{1}{2}\left[\int_{q \le p} q + \int_{q > p} p\right] =$$

$$\tfrac{1}{2}\left[\int_{q \le p} q + \int_{q/p > 1+\delta} p + \int_{1 < q/p < 1+\delta} p\right] \ge$$

$$\tfrac{1}{2}\left[\int_{q \le p} q + \int_{1 < q/p \le 1+\delta} (p/q)\, q\right] >$$

$$\tfrac{1}{2}\left[\frac{1}{1+\delta}\int_{q \le p} q + \frac{1}{1+\delta}\left(\int_{q/p > 1} q - \int_{q/p > 1+\delta} q\right)\right]$$

$\ge \dfrac{1-\delta}{2(1+\delta)}$. Hence if P^i_{error} does not approach $\frac{1}{2}$, such a δ exits. But then the divergence $J^i(p,q) =$

$$\int_{p \ge q} \ln(p/q)(p-q) + \int_{q > p} \ln(q/p)(q-p)$$

$$\geq \left[\ln(1+\delta)\right]\left[\left(1-\frac{1}{1+\delta}\right)\delta\right] = \frac{\delta^2 \ln(1+\delta)}{1+\delta} > 0 .$$

CASE (2) $\quad J^i \longrightarrow J \neq 0$

Let's rewrite

$$h^i(X) = \tfrac{1}{2}\sum_{1}^{n_i}\left[(x_\ell^i)^2\left[1-1/\lambda_\ell^i\right] - 2x_\ell^i\, d_\ell^i\right] + K_i$$

where we reorder the d_ℓ^i such that

$$d_\ell^i \geq d_{\ell+1}^i \quad .$$

Subcase (a) $\quad \sup_i\left(\sum_{1}^{n_i}(d_\ell^i)^2\right) = +\infty \quad .$

Clearly from (5) $J = +\infty$. Consider the (sub-optimal) discriminants $g^i = \sum_{1}^{n_i} x_\ell^i\, d_\ell^i$. These are normally distributed with means, $\sum_{1}^{n_i}(d_\ell^i)^2$ and 0, and standard deviations, $\sqrt{\sum_{1}^{n_i}(d_\ell^i)^2}$ and $\sqrt{\sum_{1}^{n_i}\lambda_\ell^i\,(d_\ell^i)^2}$. One can then find arbitrarily large i for which g^i has arbitrarily small error probability. Since h^i is optimal, it has arbitrarily small error for these i and hence, $P^i_{error} \longrightarrow 0$.

Subcase (b) $\sup_i \left(\sum_1^{n_i} (d_\ell^i)^2 \right) < +\infty$.

We first note that $\mathrm{VAr}(h^i) \longrightarrow J \neq 0$ under either hypothesis.

Let us rewrite $h^i = \frac{1}{2} \sum_1^{\overline{n_i}} \left[(x_\ell^i)^2 \left[1-1/\lambda_\ell^i\right] - 2x_\ell^i\, d_\ell^i \right]$

$+ \frac{1}{2} \sum_{\overline{n_i}+1}^{n_i} \left[(x_\ell^i)^2 \left[1-1/\lambda_\ell^i\right] - 2x_\ell^i\, d_\ell^i \right] + K_i = F_1^i + F_2^i + K_i$ with

$\overline{n_i}$ chosen such that $\overline{n_i} \longrightarrow \infty$ but that $\sum_1^{\overline{n_i}} |1-1/\lambda_\ell^i| \longrightarrow 0$. We may now apply a central limit theorem, for instance Corollary 4.2 on page 232 of [1]: F or any $\beta > 0$, either F_2^i has variance $<\beta$, or F_2^i becomes normal in distribution for large i. This follows from the central limit theorem for arrays mentioned above, provided the variances of the terms in the summand of F_2^i become arbitrarily small and this follows if $\sup_i \left\{ d_\ell^i ; \ell > \overline{n_i} \right\} \longrightarrow 0$. But if this were not the case, $d_{\overline{n_i}+1}^i \geq \gamma > 0$ for infinitely many i and hence, since $\overline{n_i} \longrightarrow \infty$, $\sum_1^{n_i} (d_\ell^i)^2 \geq \overline{n_i}\, \gamma^2$ contradicts our initial assumption. Further, F_1^i either has variance $<\beta$ or approaches a normal random variable in distribution since its linear part is normal and its nonlinear part has variance approaching 0. Since β was arbitrary, $J > 0$, and F_1^i is independent of F_2^i ; h^i approaches a normal random variable in distribution and we obtain the asymptotic error

formula (2).

Finally we note that, in (2), we could replace J by 8B where B is the Bhattacharyya distance. This follows from the simply verified fact that $\frac{8B}{J} \longrightarrow 1$ as all $\lambda_\ell \longrightarrow 1$.

REFERENCES

[1]. L. H. Y. Chen, "Two Central Limit Problems for Dependent Random Variables," Z. Wahrscheinlichkeitstheorie verw. Gebeite 43 , 223-243 (1978).

[2]. K. Fukunaga, Introduction to Statistical Pattern Recognition (Academic Press, New York, 1972).

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER ESD-TR-80-193	2. GOVT ACCESSION NO. AD-A094 066	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Some Remarks on the Gaussian Discriminant		5. TYPE OF REPORT & PERIOD COVERED Technical Note
		6. PERFORMING ORG. REPORT NUMBER Technical Note 1980-47
7. AUTHOR(s) Lee K. Jones		8. CONTRACT OR GRANT NUMBER(s) F19628-80-C-0002
9. PERFORMING ORGANIZATION NAME AND ADDRESS Lincoln Laboratory, M.I.T. P.O. Box 73 Lexington, MA 02173		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS Program Element No. 63311F Project No. 627A
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Systems Command, USAF Andrews AFB Washington, DC 20331		12. REPORT DATE 3 Oct[illegible]980
		13. NUMBER OF PAGES 16
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) Electronic Systems Division Hanscom AFB Bedford, MA 01731		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

None

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Gaussian discriminant function

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

We comment on the performance of the Gaussian discriminant function with (possibly) non-Gaussian underlying distributions. An asymptotic expression for the probability of error for the Gaussian case is given with a formal convergence proof.

DD FORM 1473 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

END

DATE
FILMED

2 - 81

DTIC